REMARKS

Favorable reconsideration of this application, in light of the following discussion, is respectfully requested.

Claims 1-24 are currently pending, with Claims 1-8 and 15-20 withdrawn as directed to a nonelected invention. No claims have been amended herewith.

In the outstanding Office Action, Claims 9-14, 21, and 23 were rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,684,715 to Palmer (hereinafter "the '715 patent") in view of U.S. Patent No. 6,404,901 to Itokawa (hereinafter "the '901 patent"), further in view of U.S. Patent No. 6,445,409 B1 to Ito et al. (hereinafter "the '409 patent"); and Claims 22 and 24 were rejected under 35 U.S.C. § 103(a) as being unpatentable over the '715, '901, and '409 patents, further in view of U.S. Patent No. 6,278,466 to Chen (hereinafter "the '466 patent").

The '409 patent, which is asserted against all pending claims in the present application, has a filing date of June 28, 1999. However, the '409 patent was a *continuation-in-part* of Application Serial No. 09/078,521, filed on May 14, 1998. Thus, any disclosure added to the '409 patent after the filing of the '521 parent application has an effective reference date under 35 U.S.C. § 102(e) of June 28, 1999. Further, Applicants note that the Office Action relies on the '409 patent to disclose that the feature data of a predetermined object includes color information of an area of the predetermined object, as recited in independent Claim 9. However, Applicants respectfully submit that this disclosure was added to the '409 patent application and is not found in the '521 parent application.[1]

The actual U.S. filing date of the present application is January 28, 2000. However, the present application claims priority under 35 U.S.C. § 119 to Japanese Patent Application

[1] See U.S. Patent No. 6,404,455 B1, which is based on the '521 parent application.

No. P11-022372, filed January 29, 1999.[2] Therefore, in order to perfect this claim for foreign priority in the present application, submitted herewith is a certified English translation of Japanese Patent Application No. JP P11-022372. Accordingly, Applicants respectfully submit that the '409 patent disclosure relied upon by the Office Action does not qualify as a *prima facie* prior art reference against the claims in the present application. Accordingly, the rejections in the outstanding Office Action should be withdrawn.

Thus, it is respectfully submitted that independent Claim 9 (and dependent Claims 10-13, 21, and 22) and independent Claim 14 (and dependent Claims 23 and 24) patentably define over any proper combination of the '715, '901, and '409 patents.

Consequently, in light of the above discussion, the outstanding grounds for rejection are believed to have been overcome. The application is believed to be in condition formal allowance. An early and favorable action to that effect is respectfully requested.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.

Kurt M. Berger

Eckhard H. Kuesters
Attorney of Record
Registration No. 28,870
Kurt M. Berger, Ph. D.
Registration No. 51,461

Customer Number
22850

Tel: (703) 413-3000
Fax: (703) 413-2220
(OSMMN 08/03)

EHK/KMB/law
I:\ATTY\KMB\00397544-AM1.DOC

[2] This claim for foreign priority was acknowledged in the outstanding Office Action.

CERTIFICATE

I, the undersigned, Tatsuo HASEDA, residing at 2-16-16, Saginomiya, Nakano-ku, Tokyo, JAPAN, hereby certify that to the best of my knowledge and belief the following is a true translation into English made by me of Japanese Patent Application Number Hei 11-22372 filed on January 29, 1999.

Date: December 16, 2003

T. Haseda

Name: Tatsuo Haseda

[Name of Document] Application for Patent

[Reference No] A009900192

[Application Date] January 29, 1999

[Destination] Commissioner, Patent Office

[International Patent Classification] G06F 15/00

[Title of the Invention] Video Information Description Method, Video Retrieval Method, and Video Retrieval Device

[No. of Claims] 17

[Inventor]

[Address] 1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki-shi, Kanagawa-ken

Research and Development Center, TOSHIBA CORPORATION

[Name] Kohji Yamamoto

[Inventor]

[Address] 1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki-shi, Kanagawa-ken

Research and Development Center, TOSHIBA CORPORATION

[Name] Osamu Hori

[Inventor]

[Address] 1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki-shi, Kanagawa-ken

Research and Development Center, TOSHIBA CORPORATION

[Name] Toshimitsu Kaneko

[Inventor]

[Address] 1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki-shi, Kanagawa-ken

Research and Development Center, TOSHIBA CORPORATION

[Name] Takeshi Mita

[Applicant for Patent]

[Identification No.] 000003078

[Name] TOSHIBA CORPORATION

[Agent]

[Identification No.] 100058479

[Patent Attorney]

[Name] Takehiko Suzue

[Telephone No.] 03-3502-3181

[Assigned Agent]

[Identification No.] 100084618

[Patent Attorney]

[Name] Sadao Muramatsu

[Assigned Agent]

[Identification No.] 100068814

[Patent Attorney]

[Name] Atsushi Tsuboi

[Assigned Agent]

[Identification No.] 100092196

[Patent Attorney]

[Name] Yoshiro Hashimoto

[Assigned Agent]

[Identification No.] 100091351

[Patent Attorney]

[Name] Satoshi Kohno

[Assigned Agent]

[Identification No.] 100088683

[Patent Attorney]

[Name] Makoto Nakamura

[Assigned Agent]

[Identification No.] 100070437

[Patent Attorney]

[Name] Shoji Kawai

[Designation of Charge]

[Ledger No. for Prepayment] 011567

[Amount of Payment] 21000

[List of Filed Document]

[Object Name] Specification 1

[Object Name] Drawings 1

[Object Name] Abstract 1

[Necessity of Confirmation] Necessary

[Name of Document] Specification

[Title of the Invention] Video Information Description Method, Video Retrieval Method, and Video Retrieval Device

[What is Claimed is]

[Claim 1] A video information description method describing characteristic amounts concerning a specific object on a screen and characteristic amounts concerning a background on said screen as video information.

[Claim 2] A video information description method describing characteristic amounts concerning a specific object on a screen, characteristic amounts concerning a background on said screen, and differences between said characteristic amounts as video information.

[Claim 3] A video information description method according to either of Claims 1 and 2, wherein said method describes at least information of the position, shape, and motion of said object as said characteristic amounts concerning said object and describes at least information of the motion of said background as said characteristic amount concerning said background.

[Claim 4] A recording medium storing said characteristic amounts concerning said object and said characteristic amounts concerning said background described by said video information description method stated in any one of Claims 1 to 3 together with video data or separately from said video data.

[Claim 5] An object detection method comprising a moving vector extraction step of extracting a moving vector of input video, an inference step of inferring a motion of a background of said video using said moving vector, and a detection step of removing said inferred motion of said background, detecting a moving vector concerning a specific object on a screen, and detecting an area of said object.

[Claim 6] An object detection method according to Claim 5, wherein said inference step approximates said motion of said background to a predetermined transformation model, infers a transformation coefficient of said transformation model from said moving vector of said video, thereby infers said motion of said background.

[Claim 7] An object detection method according to Claim 6, wherein said inference step infers said transformation coefficient of said transformation model by a low burst inferring method.

[Claim 8] An object detection method according to any one of Claims 5 to 7, wherein said inference step comprises a process of dividing said moving vectors in a screen of said video into areas according to a degree of similarity, clustering said divided areas on the basis of said degree of similarity of said moving vectors, and deciding a largest cluster area as an area of said background.

[Claim 9] An object detection method according to any

one of Claims 5 to 7, wherein said inference step comprises a process of dividing said moving vectors of a plurality of frames of said video into areas according to the degree of similarity in each frame, bringing said areas into correspondence to each other between said frames, clustering said areas in said frames on the basis of said moving vectors so as to allow said corresponded areas to belong to the same cluster, and deciding a largest cluster area as an area of said background.

[Claim 10] A video retrieval method describing characteristic amounts concerning a specific object to be retrieved on a screen and characteristic amounts concerning a background on said screen and

subtracting said characteristic amounts concerning said background from said characteristic amounts concerning said object on said screen, then comparing said differences with characteristic amounts concerning an object input from the outside, thereby retrieving, from said video to be retrieved, the same object as said object input from the outside or at least one of said frames on said screen including the same object as said object input from the outside.

[Claim 11] A video retrieval method describing characteristic amounts concerning a specific object to be retrieved on a screen, characteristic amounts concerning a background on said screen, and differences between said

respective characteristic amounts and

comparing said differences with characteristic amounts concerning an object input from the outside, thereby retrieving, from said video to be retrieved, the same object as said object input from the outside or at least one of said frames on said screen including the same object as said object input from the outside.

[Claim 11] A video retrieval method describing characteristic amounts concerning a specific object to be retrieved on a screen and characteristic amounts concerning a background on said screen and

comparing said characteristic amounts concerning said background with characteristic amounts concerning said background inputted from the outside on said screen, thereby retrieving, from said video to be retrieved, a frame using almost the same camera work as a camera work when said video input from the outside is obtained.

[Claim 13] A video retrieval method describing characteristic amounts at least including motion information concerning a specific object to be retrieved on a screen and characteristic amounts concerning a background on said screen and

comparing motion information of an object in a plurality of continuous frames on said screen with information of a series of motions of an object input from the outside, thereby retrieving, from said video to be

retrieved, the same object as said object input from the outside or at least one of said frames on said screen including the same object as said object input from the outside.

[Claim 14] A video retrieval apparatus describing characteristic amounts concerning a specific object to be retrieved on a screen and characteristic amounts concerning a background on said screen and

subtracting said characteristic amounts concerning said background from said characteristic amounts concerning said object on said screen, then comparing said differences with characteristic amounts concerning an object input from the outside, thereby retrieving, from said video to be retrieved, the same object as said object input from the outside or at least one of said frames on said screen including the same object as said object input from the outside.

[Claim 15] A video retrieval apparatus describing characteristic amounts concerning a specific object to be retrieved on a screen, characteristic amounts concerning a background on said screen, and differences between said respective characteristic amounts and

comparing said differences with characteristic amounts concerning an object input from the outside, thereby retrieving, from said video to be retrieved, the same object as said object input from the outside or at

least one of said frames on said screen including the same object as said object input from the outside.

[Claim 16] A video retrieval apparatus describing characteristic amounts concerning a specific object to be retrieved on a screen and characteristic amounts concerning a background on said screen and

comparing said characteristic amounts concerning said background with characteristic amounts concerning said background inputted from the outside on said screen, thereby retrieving, from said video to be retrieved, a frame using almost the same camera work as a camera work when said video input from the outside is obtained.

[Claim 17] A video retrieval apparatus describing characteristic amounts at least including motion information concerning a specific object to be retrieved on a screen and characteristic amounts concerning a background on said screen and

comparing motion information of an object in a plurality of continuous frames on said screen with information of a series of motions of an object input from the outside, thereby retrieving, from said video to be retrieved, the same object as said object input from the outside or at least one of said frames on said screen including the same object as said object input from the outside.

[Detailed Description of the Invention]

[0001]

[Background of the Invention]

[Field of the Invention]

The present invention relates to a video information description method taking notice of an object on the screen and a video retrieval method and a video retrieval device for retrieving a specific object and a frame including it using the above method.

[0002]

[Description of the Related Art]

Due to multi-channel of digital satellite broadcast and broadcast in correspondence with the spread of table television, video information obtainable by users keeps on increasing. On the other hand, due to progress of the computer technique and practical realization of bulk recording media represented by a DVD, it is becoming easy to store a large amount of video information as digital information and handle them by computers.

[0003]

When a user, to actually use video information, is to efficiently access target video among such a large amount of video information, an effective video retrieval art is necessary. As such a video retrieval art, a method for adding any information to an object on the screen, retrieving video including an object satisfying the information required by the user, and allowing him to view

it is under consideration. To add any information to an object on the screen, a process of extracting an object from video is necessary. However, it is not realistic to manually extract an object from video information keeping on increasing.

[0004]

Regarding an automatic object detection art, for example, in the document, 'Yoneyama, Nakajima, Yanagihara, and Sugano, "Detection of Moving Object from MPEG Video Stream", Shin-Gaku Paper, Vol. J81-D-II, No. 8, pp. 1776-1786, Aug., 1998', a method of detecting an object from video with a static background is proposed. However, in this method, the static background is a prerequisite and when the background moves, no object can be detected easily.

[0005]

Namely, even if the shape of an object is given beforehand, when the motion of the background is not known, to retrieve the object using the motion thereof, the motion is adversely affected by the camera work, thus the object cannot be retrieved using a precise motion. For example, when an object moving to the left is chased and imaged, the object almost stands still in the screen and the background relatively moves to the right. As a result, the video including the object moving to the left in the screen cannot be retrieved.

[0006]

[Problems to be Solved by the Invention]

As mentioned above, the conventional video retrieval art cannot retrieve an object with a moving background, so that a problem arises that video including such an object cannot be retrieved.

[0007]

An object of the present invention is to provide a video information description method capable of retrieving video including an object with a moving background.

[0008]

Another object of the present invention is to provide an object detection method capable of detecting an object with a moving background.

[0009]

Still another object of the present invention is to provide a video retrieval method and a video retrieval device capable of variously retrieving video including an object with a moving background.

[0010]

[Means for Solving the Problems]

To solve the above problems, a video information description method relating to the present invention is basically characterized in that it describes characteristic amounts concerning a specific object on the screen and characteristic amounts concerning the

background on the screen as video information.

[0011]

Further, another video information description method relating to the present invention is characterized in that it describes, in addition to characteristic amounts concerning a specific object on the screen and characteristic amounts concerning the background on the screen, moreover the differences between the two characteristic amounts as video information. Further, it may describe the differences between characteristic amounts concerning a specific object on the screen and characteristic amounts concerning the background on the screen and the characteristic amounts concerning the background as video information.

[0012]

Here, it is preferable to describe at least information of the position, shape, and motion of the object as characteristic amounts concerning a specific object and at least information of the motion of the background as a characteristic amount concerning the background.

[0013]

Further, according to the present invention, a recording medium in which the characteristic amounts concerning an object, the characteristic amounts concerning the background, and moreover the differences between the two characteristic amounts which are described

in this way are stored together with or separately from video data is provided.

[0014]

The object detection method relating to the present invention is characterized in that it has a moving vector extraction step of extracting the moving vector of input video, an inference step of inferring the motion of the video background using the extracted moving vector, and a detection step of removing the inferred motion of the background, extracting the moving vector concerning a specific object on the screen, and detecting the area of the object.

[0015]

Here, the inference step of inferring the motion of the background is characterized in that it approximates the motion of the background to a predetermined transformation model (for example, the affine transformation or perspective transformation), infers the transformation coefficient of the transformation model from the moving vector of the video, thereby infers the motion of the background. The transformation coefficient of the transformation model is inferred, for example, by the low burst inferring method.

[0016]

The inference step of inferring the motion of the background may include a process of dividing the moving

vectors in the video screen into areas according to the degree of similarity, clustering the divided areas on the basis of the degree of similarity of the moving vectors, and deciding the largest cluster area as an area of the background.

[0017]

Further, the inference step of inferring the motion of the background may include a process of dividing the moving vectors of a plurality of frames of video into areas according to the degree of similarity in each frame, bringing the areas into correspondence to each other between the frames, clustering the areas in the frames on the basis of the moving vectors so as to allow the corresponded areas to belong to the same cluster, and deciding the largest cluster area as an area of the background.

[0018]

The present invention describes a characteristic amount concerning an object on the screen, a characteristic amount concerning the background, and moreover the differences between the two characteristic amounts, thereby can variously retrieve video including an object with a moving background.

[0019]

Firstly, the present invention describes characteristic amounts concerning a specific object on the

screen which is to be retrieved and characteristic amounts concerning the background on the screen, subtracts the characteristic amounts concerning the background from the characteristic amounts concerning the specific object on the screen which is to be retrieved, then compares the differences with characteristic amounts concerning an object input from the outside, thereby can retrieve, from the video to be retrieved, the same object as the object input from the outside or at least one of the frames on the screen including the same object as the object input from the outside.

[0020]

Secondly, the present invention describes characteristic amounts concerning a specific object on the screen which is to be retrieved, characteristic amounts concerning the background on the screen, and the differences between the characteristic amounts, compares the differences with characteristic amounts concerning an object input from the outside, thereby can retrieve, from the video to be retrieved, the same object as the object input from the outside or at least one of the frames on the screen including the same object as the object input from the outside.

[0021]

Thirdly, the present invention describes characteristic amounts concerning a specific object on the

screen which is to be retrieved and characteristic amounts concerning the background on the screen, compares the characteristic amounts concerning the background with characteristic amounts concerning the background on the screen which is input from the outside, thereby can retrieve, from the video to be retrieved, a frame using almost the same camera work as the camera work when the video input from the outside is obtained.

Fourthly, the present invention describes characteristic amounts at least including motion information concerning a specific object on the screen which is to be retrieved and characteristic amounts concerning the background on the screen, compares motion information of an object in a plurality of continuous frames on the screen with information of a series of motions of an object input from the outside, thereby can retrieve, from the video to be retrieved, the same object as the object input from the outside or at least one of the frames on the screen including the same object as the object input from the outside.

[0022]

[Description of the Preferred Embodiments]

The embodiments of the present invention will be explained hereunder with reference to the accompanying drawings.

[First embodiment]

This embodiment provides broadly three functions. Firstly, the embodiment provides a function for reproducing video data and additionally a function for automatically detecting a moving object on the screen, overlapping, composing, and displaying elliptic or rectangular figures, thereby informing a user of the existence thereof.

[0023]

Secondly, the embodiment provides a function for separating characteristic amounts of position, size, and motion of a detected object from characteristic amounts concerning the background and describing them in an external file as display data.

[0024]

Thirdly, the embodiment provides a function for comparing data of characteristic amounts concerning a detected object or display data of characteristic amounts described in an external file beforehand with data of characteristic amounts of a retrieval object given externally as an object to be retrieved, presenting a corresponding object to a user, thereby retrieving the object on the screen.

[0025]

Fig. 1 shows the constitution and procedure of the video retrieval system relating to this embodiment as a flow chart.

[0026]

Firstly, the video retrieval system inputs original video data 100 reproduced from a medium such as a DVD (Step 101) and detects a specific object on the screen from original video data 101 by a method which will be explained later in detail (Step 102). In this case, as described later, the video retrieval system also detects information concerning the background on the screen. The system composes the detected object with an elliptic or rectangular figure generated so as to surround it and outputs it as object detection result display data 104 (Step 103).

[0027]

On the other hand, the video retrieval system performs a characteristic amount data generation process for describing characteristic amount data indicating characteristic amounts concerning an object detected at Step 102 such as the position, shape (including the size), and motion and characteristic amount data indicating characteristic amounts concerning the background such as the motion of the background (Step 105) and moreover performs a process of outputting characteristic amount data 107 concerning the generated object and background to the outside and describing as display data (Step 106).

[0028]

Further, at Steps 105 and 106, characteristic amount data concerning an object and characteristic amount data

concerning the background may be described. However, furthermore, data of the differences between the two characteristics may be generated and described or according to circumstances, the difference data and characteristic amount data concerning the background or the difference data and characteristic amount data concerning the object may be generated and described.

[0029]

The description process at Step 106 is concretely a process of storing (recording) and displaying characteristic amount data 107 in various recording media or memories. A recording medium storing characteristic amount data 107 may be a medium such as the DVD storing original video data 100 or may be a recording medium different from it.

[0030]

Next, the video retrieval system, to retrieve the object, decides the degree of similarity between the characteristic amount data concerning the object generated at Step 105 and retrieval object characteristic amount data 110 inputted at Step 109 and furthermore, performs a composition display process for composing and displaying the similarity decision result as object retrieval result display data 112 (Step 111). Retrieval object characteristic amount data 110 is data displaying characteristic amounts such as the position, shape

(including the size), and motion of an object to be retrieved.

[0031]

Further, the series of processes can be realized by either of the software and hardware.

[0032]

Next, by referring to Fig. 2, the object retrieval process at Step 102 shown in Fig. 1 will be explained in detail.

Firstly, the video retrieval system extracts a moving vector from original video data 100 inputted (Step 201). When original video data 100 is MPEG compressed data, the system uses a moving vector obtained from picture P. In this case, moving vectors are given for each macro-block. When original video data 100 is analog data or digital data having no moving vector, the system digitizes it as required, extracts a moving vector by using an optical flow, converts it to MPEG compressed data, and then extracts a moving vector.

[0033]

The moving vector extracted like this may not be always reflected by the motion of an actual object and it is conspicuous in the peripheral part of the screen and in a part having a flat texture. Therefore, the video retrieval system performs a process of removing moving vectors with low reliability (Step 202). This process is

performed as indicated below.

[0034]

Firstly, with respect to the peripheral part of the screen, an area is decided beforehand and moving vectors included in the area are removed. On the other hand, with respect to the part having a flat texture, when original video data 100 is MPEG compressed data, using the DC component with the DCT (discrete cosine transformation) coefficient of picture I as shown in Fig. 3, as shown in Fig. 4, the group of macro-blocks in which the dispersal of the four DC components contained in one macro-block is smaller than the threshold value is assumed as a low reliable area and a moving vector whose starting point is included in the macro-block in the area is removed as a moving vector with low reliability.

[0035]

In the moving vector data obtained in this way, since the motion of the background due to the camera work is included in the object motion, to obtain a precise object motion, the motion of the background must be removed. Therefore, in this embodiment, as a transformation model for approximating to the motion of the background due to the camera work, an affine transformation model is used and a process of inferring the transformation coefficient thereof by using a moving vector, thereby inferring the motion of the background is performed (Step 203). With

respect to the process of inferring the affine transformation coefficient of the motion of the background, there are several kinds of methods available, which will be described later.

[0036]

Next, a process of converting the starting point of each moving vector by using the inferred affine transformation coefficient, subtracting the moved part thereof from the original moving vector, thereby removing the motion of the background is performed (Step 204).

[0037]

A process of dividing the moving vector data not including the motion of the background obtained in this way into areas composed of similar moving vectors is performed (Step 205). Concretely, neighboring two moving vectors are compared in the cosine (direction) and magnitude and when the differences are smaller than predetermined threshold values, the process of dividing into the same area is carried out for all the combinations of neighboring moving vectors.

[0038]

In the areas obtained in this way, small areas unsuitable for handling as an object are included, so that a decision process 306 of removing those unsuitable areas by the threshold process is performed and finally object data 307is output.

[0039]

The three methods of the affine transformation coefficient inference process at Step 203 of approximating to the motion of the background from the moving vectors will be explained below.

[0040]

<Method 1>

Method 1 infers the affine transformation coefficient by using all the moving vectors in the screen excluding the moving vectors with low reliability. The center of the i-th macro-block is assumed as y_i and the moving vector corresponding to the macro-block is assumed as v_i. At this time, the moving destination by an affine transformation deformation model at the starting point $x_i = y_i - v_i$ of the vector, assuming the affine transformation coefficient as ∂, is $r_i = x_i \partial$ and the error from the actual moving destination y_i is $e_i = r_i - y_i$. The sum total of inferred residuals is expressed by the following formula and ∂ for minimizing it must be obtained.

[0041]

[Formula 1]

$$\sum_i \Psi(e_i / \sigma_i) = \min$$

[0042]

As a method for solving such a problem, there is the method of least squares and in such a case, in Formula (1),

$\Psi(z) = z^2$ may be used. However, when the method of least squares is used, the moving vectors of the background and the moving vectors of an object are dealt with each other on the same basis, so that no affine transformation coefficient can be inferred only from the moving vectors of the background and an affine transformation coefficient including the motion of the object is obtained.

[0043]

Therefore, assuming the area occupied by the background area in the screen as 50% or more and regarding the moving vectors of the object as a disturbance, an affine transformation coefficient is inferred only from the moving vectors of the background. As a method strong for disturbance, a low burst inference method as disclosed in Document, Tohru Nakagawa and Yoshio Koyanagi, "Experimental Data Analysis by Method of Least Squares", Publication Association, Tokyo University is used. Here, particularly, M inference by the Biweight method which is a low burst inference method is used. The Biweight method lowers the weight which is an element causing a large error, thereby enables to be hardly affected by disturbance. Concretely, in $\Psi(z)$ of Formula (1), Formula (2) using a weight of w is used. It is said that a constant of c is preferably selected from 5 to 9.

[0044]

[Formula 2]

$$\Psi(z) = \int w_j \, z \, dz$$

$$w_j = \begin{cases} (1 - (z_j / c)^2)^2, & |z_j| < c \\ 0, & \text{otherwise} \end{cases}$$

[0045]

<Method 2>

In the affine transformation coefficient inference process of approximating to the motion of the background, the procedure when Method 2 is used will be explained by referring to the flow chart shown in Fig. 5.

[0046]

Firstly, for moving vector data 500 after the process of removing vectors with low reliability is performed, using the same process as that of Method 1, the process of dividing neighboring moving vectors into similar areas is performed (Step 501). However, unlike Method 1, the process of removing the motion of the background is not performed at this time.

[0047]

Next, the affine transformation coefficient inference process when the affine transformation model approximates to the area motion is performed by the moving vectors included in the divided areas (Step 502). For the inference process at this time, the low burst inference method similar to that of Method 1 is used.

[0048]

Next, the clustering process is performed for the

divided areas (Step 503). For it, a table composed of combinations of all areas is prepared and the distances between the areas are obtained from the affine transformation coefficient. Here, the Euclid distance of a coefficient of 6 of the affine transformation model is used. However, other distances may be used. Next, two areas having a shortest Euclid distance are united, and a new affine transformation coefficient is obtained for the united area, and the two united areas are deleted from the table, and the united area is added, thus the table is updated. This process is repeated until the inter-area distance becomes larger than a predetermined threshold value or until the areas are reduced to one area.

[0049]

A process of deciding, among the areas clustered in this way, the area having a largest cluster as an area of the background is performed (Step 504) and the affine transformation coefficient of the area is output as an affine transformation coefficient 505 of the motion of the background.

[0050]

<Method 3>

In the affine transformation coefficient inference process of approximating to the motion of the background, the procedure when Method 3 is used will be explained by referring to the flow chart shown in Fig. 6.

[0051]

Firstly, a plurality of frames are read at a time and a process of dividing neighboring moving vectors into similar areas by using the same process as that of Method 2 for the frames (Step 601).

[0052]

Next, an inference process of a transformation coefficient when the affine transformation model approximates to the motion of each area is performed (Step 602), and furthermore, on the basis of the position of each area, moving vector data, and transformation coefficient, a process of obtaining an area corresponding to the frames is performed (Step 603), and then the area in each frame is clustered by the same clustering process as that of Method 2 (Step 604).

[0053]

When an area brought into correspondence by the inter-frame corresponding process is clustered to another cluster, the result clustered to most clusters is a right answer and a correction process of moving an area clustered to another cluster is performed (Step 605).

[0054]

Finally, a decision process of deciding an area having a largest area among the plurality of frames as a background is performed (Step 606) and transformation coefficient 507 of the background of each frame is obtained. Method 3 has

an advantage that even when the background area temporarily becomes smaller than other areas in a specific frame, the transformation coefficient can be inferred correctly.

[0055]

In the above example, for the transformation model used in the process of inferring the motion of the background, the affine transformation is used. However, other transformation models such as perspective transformation may be used.

[0056]

Next, by referring to Fig. 7, the data representation used in the description process of the characteristic amount data concerning the object and background at Step 106 shown in Fig. 1 will be explained. Here, as shown in Fig. 7(a), as an example, display data 700 of the three objects included in video 705 in the 1000th frame is displayed. Display data 700 is composed of data of frame information 701 indicating the corresponding frame in video stream 706 of the original video data, of characteristic amount 703 concerning the object, and of characteristic amount 704 concerning the background and is managed by the list structure using pointer 702 to the next display data.

[0057]

Characteristic amount 703 concerning the object includes at least information of the position, shape (including the magnitude), and motion of the object and

is concretely composed of, for example, various characteristic amounts as shown in Fig. 7(b). In this example, characteristic amounts 703 concerning the object are composed of "position", "outline" which is a shape, "affine transformation coefficient" which is information of the motion, "average and direction of moving vector", and moreover "color histogram".

[0058]

Here, the outline of the object may be approximated by a simple figure such as an ellipse or a rectangle. The affine transformation coefficient, as mentioned above, is a coefficient inferred when the motion of the object approximates to the affine transformation model. The average of moving vectors is a mean value of the magnitudes of the moving vectors in the object. Further, when object color information can be obtained, the color histogram of the object area can be used as a characteristic amount. With respect to the motion of the object, either of the motion with the background motion removed and the motion with the background motion unremoved may be recorded.

[0059]

When there are a plurality of objects as in this example, it is desirable to assign an individual ID No. to characteristic amounts 703 of each object and manage them, for example, by an expandable list structure as shown in Fig. 7(a). By use of such a list structure, the object

characteristic amounts can be easily added or deleted.

[0060]

Characteristic amounts 704 concerning the background, in the same way as with characteristic amount 703 concerning objects, are composed of various characteristic amounts as shown in Fig. 7(c), for example, "affine transformation coefficient", "average and direction of moving vectors", "camera work kind", and "color histogram". Camera work kind is referred to as typical camera work kind used for panning or zooming.

[0061]

Next, by referring to the flow chart shown in Fig. 8, the similarity decision process at Step 108 shown in Fig. 1 will be explained.

The similarity decision process is performed by comparing characteristic amount data 800 concerning each object included in the original video data with characteristic amount data 804 sequentially inputted from the outside. Characteristic amount data 804 inputted from the outside may be given by a numerical value as direct data or may be given as characteristic amount data by extracting characteristic amounts from video.

[0062]

When an object has a plurality of kinds of characteristic amounts, for each characteristic amount, the degree of similarity is sequentially obtained by the

similarity decision process (Step 803).

[0063]

For comparison of characteristic amount data 800 included in original video data 800 with characteristic amount data 804 inputted from the outside, an appropriate method is used on the basis of the characteristic amount kind. For example, when the characteristic amount is a color histogram, the use of the difference between the elements of the histogram may be considered. When objects to be compared have different kinds of characteristic amounts, only coincident characteristic amounts may be compared.

[0064]

When it is decided that at Steps 801 and 802, all the characteristic amounts data of all the objects are retrieved, a retrieval result display process is performed for information of the corresponding objects (Step 805) and the processing ends.

[0065]

Comparison of the motion of the objects may be made by removing the motion of the background by using the characteristic amount data concerning the background. By referring to Fig. 9, the retrieval effect by separation of the motion of the background will be explained.

[0066]

As shown in Fig. 9, original video data 901 is obtained

by photographing an object moving to the right by moving a camera so as to chase it, though on the screen, it is apparently seen as if the object stands still and the background moves to the right. When data of object 905 is input from the outside to retrieve the object moving to the left, the object of video data 901 stands still, so that the characteristic amounts do not coincide with each other and cannot be retrieved.

[0067]

However, when the characteristic amounts concerning the object and the characteristic amounts concerning the background are described according to the present invention, by process 902 of separating moving background 904 by the camera work using the motion of the background, object 903 accompanied by the original motion to the left of the object can be detected. Namely, in process 902, the differences between the characteristic amounts concerning the object and the characteristic amounts concerning the background are obtained, thus only object 903 is detected.

[0068]

Therefore, by comparison of detected object 902 with object 905 inputted from the outside, an object identical with object 905 inputted from the outside can be retrieved from input video data 901 and a video frame including an object identical with object 905 inputted from the outside can be retrieved from original video data 901. In this

case, when the differential data is described as mentioned above, process 902 is not necessary.

[0069]

Further, when the characteristic amounts concerning the object and the characteristic amounts concerning the background are described according to the present invention, as shown in Fig. 10, the video of the camera work coincident with the camera work inputted from the outside can be retrieved. Namely, as shown in Fig. 10, by process 1002 of separating object 1003 from original video data 1001, only background 1004 moving by the camera work is detected. And, by comparison of detected background 1004 with background 1005 moving by the camera work inputted from the outside, a video frame using a camera work coincident with the camera work inputted from the outside is retrieved from original video data 1001.

[0070]

In this case, when the differential data is described as mentioned above, process 1002 is not necessary.

[0071]

[Second embodiment]

Next, by referring to the flow chart shown in Fig. 11, the second embodiment of the present invention will be explained.

In this embodiment, in place of detection and description of the objects in the first embodiment,

original video data 1100 added with pre-analyzed characteristic amount data is input (Step 1101) and the characteristic amount data concerning the objects is separated and extracted from original video data 1100 (Step 1102).

[0072]

Then, in the same way as with the first embodiment, a similarity decision process between the characteristic amount data concerning the original video data extracted at Step 1102 and retrieval characteristic amount data 1110 inputted at Step 1109 is performed (Step 1108) and a composition display process of composing and displaying the result thereof as object retrieval result display data 1112 is performed (Step 1111).

[0073]

Further, the series of processes can be realized by either of the software and hardware.

[0074]

[Third embodiment]

Next, by referring to the flow chart shown in Fig. 12, the third embodiment of the present invention will be explained.

In this embodiment, to compare a series of motions inputted from the outside with display data extending over a plurality of frames and enable retrieval of an object by motion in time series, a process of bringing the same

objects into correspondence to each other among the objects included in a plurality of continuous display data is performed (Step 1202). On the other hand, a sampling process of extracting motion data at the same interval as that of display data 1201 from motion data 1203 inputted from the outside is performed (Step 1204).

[0075]

And, the display data corresponding to each other and the sampled externally-input motion data are compared (Step 1105) and the video including coincident objects is displayed as a retrieval result (Step 1106).

[0076]

By referring to Fig. 13, a corresponding process of the objects included in continuous display data 1201 at Step 1202 shown in Fig. 12 will be explained.

Using the characteristic amounts (position and motion) concerning object 1301 included in the N-th display data, expected position 1302 of the object in the "N+1"-th display data is obtained. And, object 1303 included in the "N+1"-th display data existing at a position closest to expected position 1302 is assumed as an object corresponding to object 1310.

[0077]

By referring to Fig. 14, a sampling process of motion data 1203 inputted from the outside at Step 1204 shown in Fig. 12 will be explained.

Motion data 1401 (same as 1203) inputted from the outside is continuous motion data, so that it cannot be compared, when it is left alone, with the display data which is discrete data added every several frames. Therefore, motion data 1401 is sampled at the frame interval of the display data and sampled motion data 1402 is compared with the display data.

[0078]

Further, the series of processes can be realized by either of the software and hardware.

[0079]

[Effects of the Invention]

As explained above, according to the present invention, the characteristic amounts concerning an object and the characteristic amounts concerning the background are described, thus the motion of the background is removed and the object can be retrieved by the original motion of the object.

[0080]

Further, video retrieval suited to the purpose of an individual user can be carried out easily by automatically detecting an object from a large amount of stored video data without assistance and extracting the characteristic amounts thereof and retrieving an object coinciding with separate characteristic amounts inputted from the outside or retrieving a frame including the same motion of the

background as that of the background accompanied by the motion by the camera work inputted from the outside.

[0081]

Furthermore, characteristic amounts detected beforehand are described, thus there is no need to perform a characteristic amount extraction process every retrieval, and high-speed retrieval can be carried out, and even if the user side has no object detection function, the aforementioned retrieval can be carried out.

[Brief Description of the Drawings]

Fig. 1 is a flow chart showing the basic procedure of the video retrieval system relating to the first embodiment of the present invention.

Fig. 2 is a flow chart showing the procedure of object detection in the first embodiment.

Fig. 3 is a drawing showing the relationship between picture I and picture P of the MPEG stream used for object detection in the first embodiment.

Fig. 4 is a drawing for explaining removal of low-reliable vectors in object detection in the first embodiment.

Fig. 5 is a flow chart for explaining a method for obtaining a transformation coefficient of the background area in the first embodiment.

Fig. 6 is a flow chart for explaining another method for obtaining a transformation coefficient of the

background area in the first embodiment.

Fig. 7 is a drawing showing the structure of the characteristic amount data used in the object description process in the first embodiment.

Fig. 8 is a flow chart showing the procedure of object retrieval in the first embodiment.

Fig. 9 is a drawing showing removal of the camera work in the object retrieval process in the first embodiment.

Fig. 10 is a drawing showing retrieval of a frame using the same camera work as the input camera work in the object retrieval process in the first embodiment.

Fig. 11 is a flow chart showing the basic procedure of the video retrieval system relating to the second embodiment of the present invention.

Fig. 12 is a flow chart showing the basic procedure of the video retrieval system relating to the third embodiment of the present invention.

Fig. 13 is a drawing for explaining correspondence of objects between the continuous display data in the third embodiment.

Fig. 14 is a drawing for explaining the sampling method of the motion data inputted from the outside in the third embodiment.

[Description of Numerals]

700: Display data

703: Characteristic amounts concerning an object

704: Characteristic amounts concerning the background

706: Video stream

[Name of Document] Abstract

[Abstract]

[Problem] The present invention provides a video information description method enabling retrieval of video including an object having a moving background.

[Solving Means] Characteristic amounts 703 including information of the position, shape, and motion of an object and information of the motion of the background are described from the original video as display data.

[Selected Drawing] Fig. 7

Name of Document: Drawing

Fig. 1

100 Original video data

101 Original video data input process

102 Object detection process

103 Detection result composition display process

104 Object detection result display data

105 Characteristic amount data generation process of the object and background

106 Characteristic amount data description process of the object and background

107 Characteristic amount data of the object and background

108 Similarity decision process of the object characteristic amounts and retrieval object characteristic amounts

109 Retrieval object characteristic amount data input process

110 Retrieval object characteristic amount data

111 Detection result composition display process

112 Object retrieval result data

Fig. 2

100 Original video data

201 Original video moving vector extraction process

202 Low-reliable moving vector removal process

203 Transformation coefficient inference process of the

background from the moving vector

204 Removal process of the moving vector of the background from the moving vector of the original video

205 Moving vector area division process

206 Object decision process of the divided areas

207 Object data

Fig. 3

1 Motion correction

2 DC component of DCT coefficient

3 Moving vector

Fig. 4

4 Small dispersal of DC component → low reliable area

5 Low reliable moving vector → removal

6 Macro-block

Fig. 5

500 Moving vector data

501 Moving vector area division process

502 Transformation coefficient inference process of each area

503 Clustering process of an area having a similar transformation coefficient

504 Background area decision process

505 Background area transformation coefficient

Fig. 6

600 Moving vector data of a plurality of frames

601 Moving vector area division process in each frame

602 Transformation coefficient inference process of each area

604 Clustering process of an area having a similar transformation coefficient

603 Area corresponding process between frames

605 Clustering correction process

606 Background area decision process

607 Background area transformation coefficient

Fig. 7

700 1000th frame display data

701 Frame No. (= 1000)

702 Pointer to next display data

704 Background characteristics

703 ID (= 1) | object characteristic amounts

a ID (= 2) | object characteristic amounts

b ID (= 3) | object characteristic amounts

c List structure

705 1000th frame video

d Object

e Background

f Video stream

g Display data

h Display data

i Frame No.

703 Object characteristic amounts (example)

- Position

- Outline (ellipse, approximate to rectangle)
- Affine transformation coefficient
- Mean and direction of moving vectors
- Color histogram

others

704 Background characteristic amounts (example)

- Affine transformation coefficient
- Mean and direction of moving vectors
- Camera work kind
- Color histogram

others

Fig. 8

800 Original video characteristic amount data

801 Any unprocessed object?

a No

805 Retrieval result display process

b Yes

c No

802 Any unprocessed characteristic amount data?

804 Characteristic amount data inputted from the outside

d Yes

803 Similarity decision process

Fig. 9

901 Original video data

a Object

904 Background

902 Background separation

903 Object

b Comparison (retrieval)

905 External input object

Fig. 10

1001 Original video data

a Object

1003 Object

1002 Background separation

1004 Background

b Comparison (retrieval)

1005 Background of external input camera work

Fig. 11

1100 Original video data with characteristic amount data

1101 Original video data input process

1102 Object characteristic amount data extraction process

1108 Similarity decision process of the characteristic amounts of original video data and retrieval object characteristic amounts

1109 Retrieval object characteristic amount data input process

1110 Retrieval object characteristic amount data

1111 Detection result composition display process

1112 Object retrieval result data

Fig. 12

1201 Display data of a plurality of frames

1202 Same object corresponding process between frames

1203 Motion data inputted from the outside

1205 Motion comparison process between display data and external input data

1204 Sampling process of external input motion data

1206 Retrieval result display process

Fig. 13

1301 Object of N-th display data

1303 Object of "N+1"-th display data

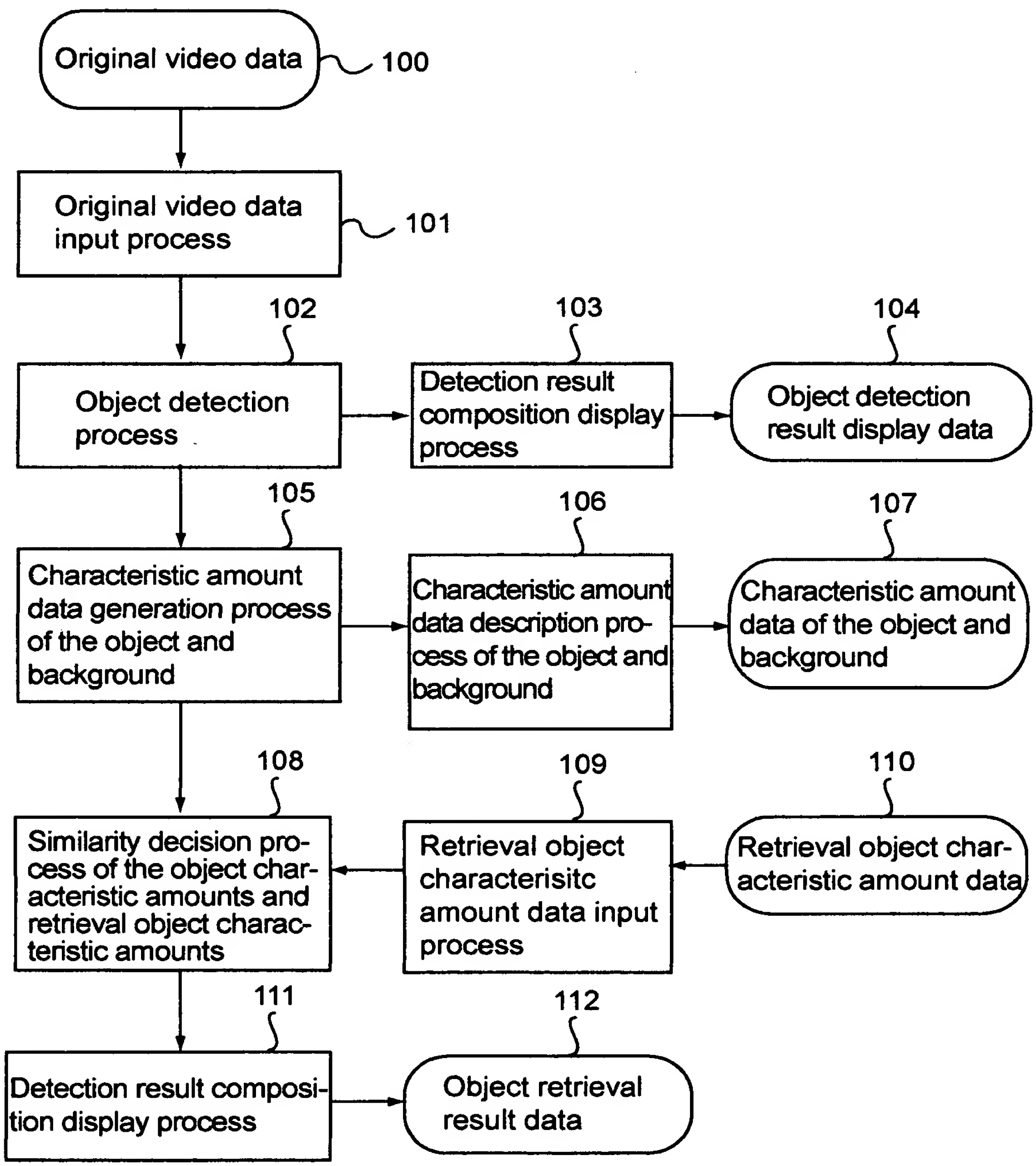
1302 Expected position

Fig. 14

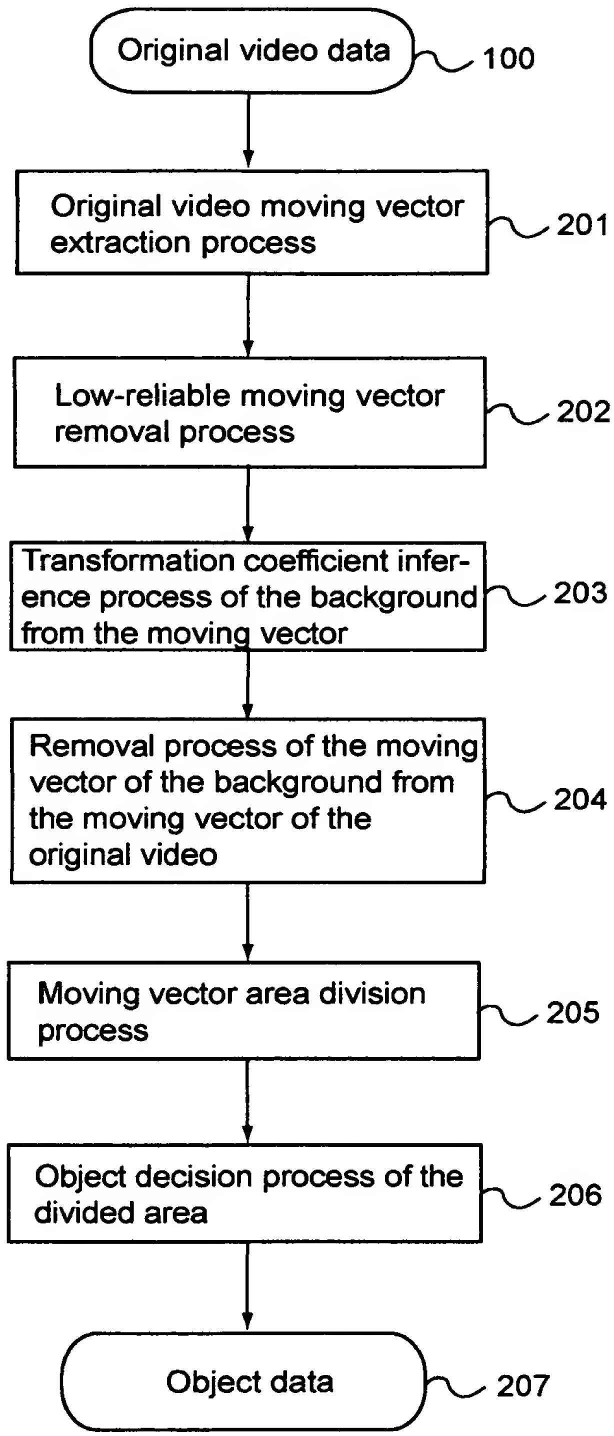
1402 Motion data sampled at frame interval of display data

1401 Inputted motion data

[FIG. 1]

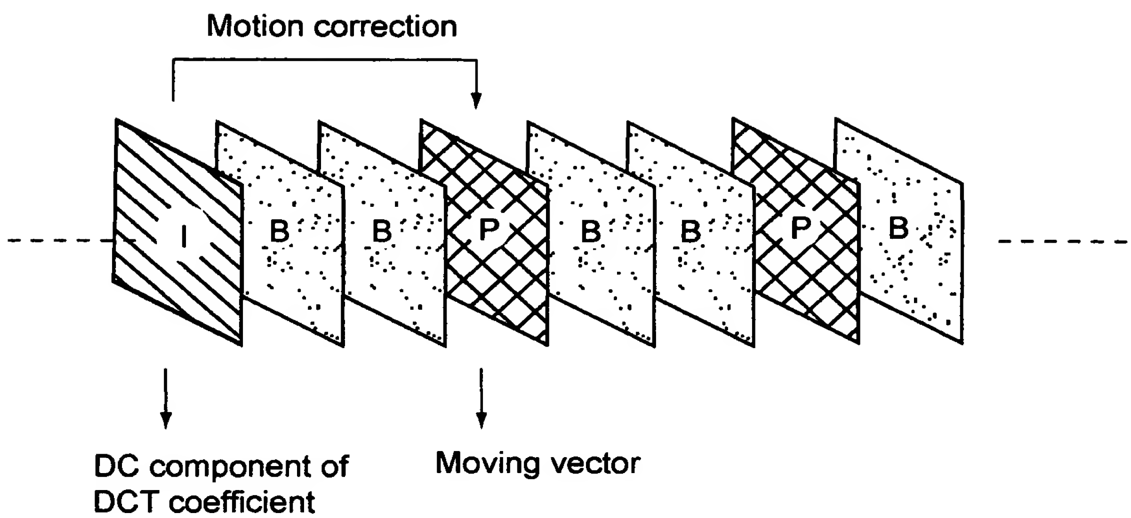

Original video data
100
Original video data input process
101
102
Object detection process
103
Detection result composition display process
104
Object detection result display data
105
Characteristic amount data generation process of the object and background
106
Characteristic amount data description process of the object and background
107
Characteristic amount data of the object and background
108
Similarity decision process of the object characteristic amounts and retrieval object characteristic amounts
109
Retrieval object characterisitc amount data input process
110
Retrieval object characteristic amount data
111
Detection result composition display process
112
Object retrieval result data

[FIG. 2]

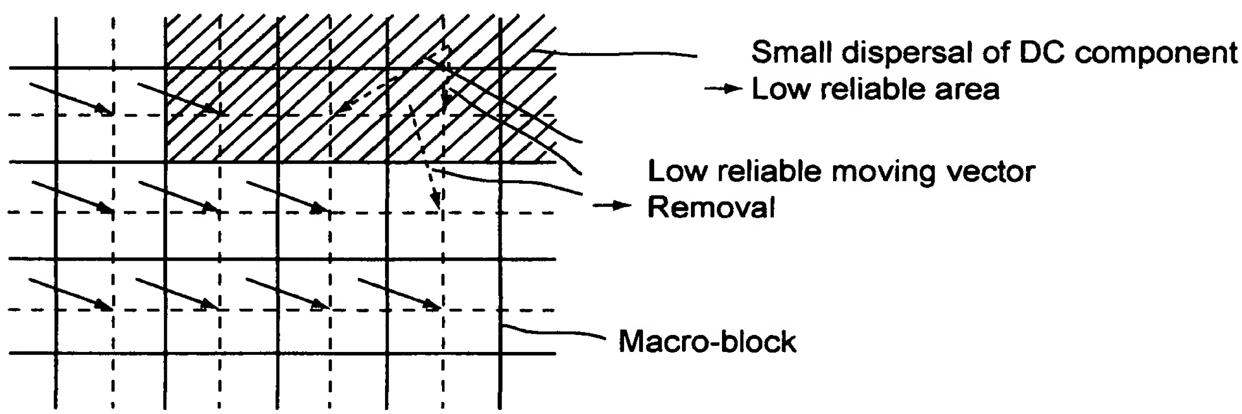


Original video data
100
Original video moving vector extraction process
201
Low-reliable moving vector removal process
202
Transformation coefficient inference process of the background from the moving vector
203
Removal process of the moving vector of the background from the moving vector of the original video
204
Moving vector area division process
205
Object decision process of the divided area
206
Object data
207

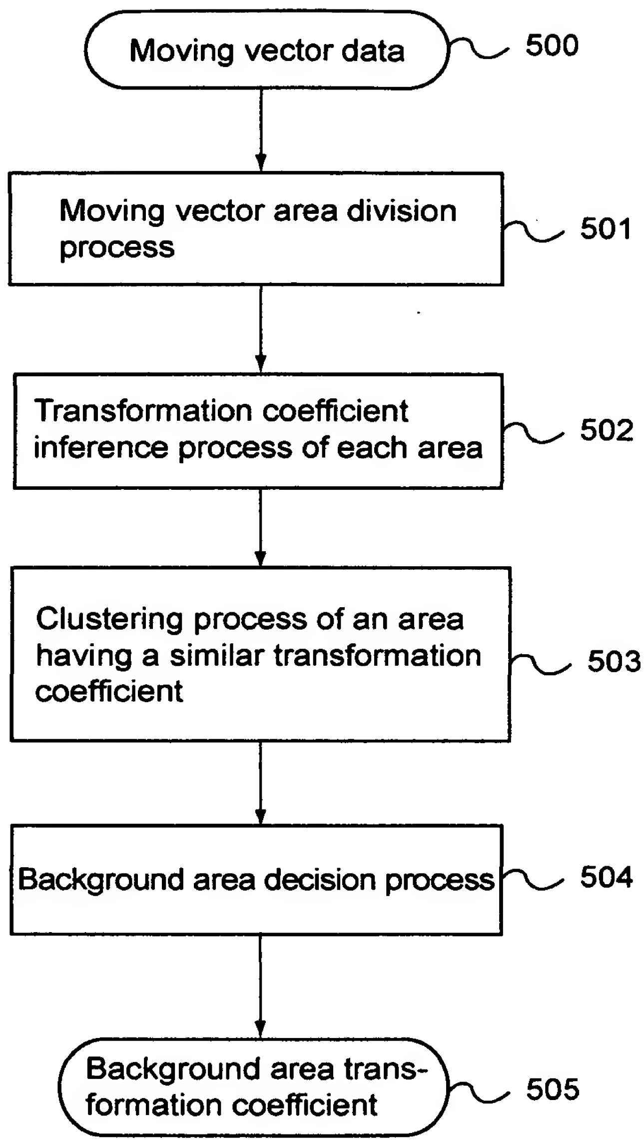
[FIG. 3]


Motion correction
I
B
B
P
B
B
P
B
DC component of DCT coefficient
Moving vector

[FIG. 4]

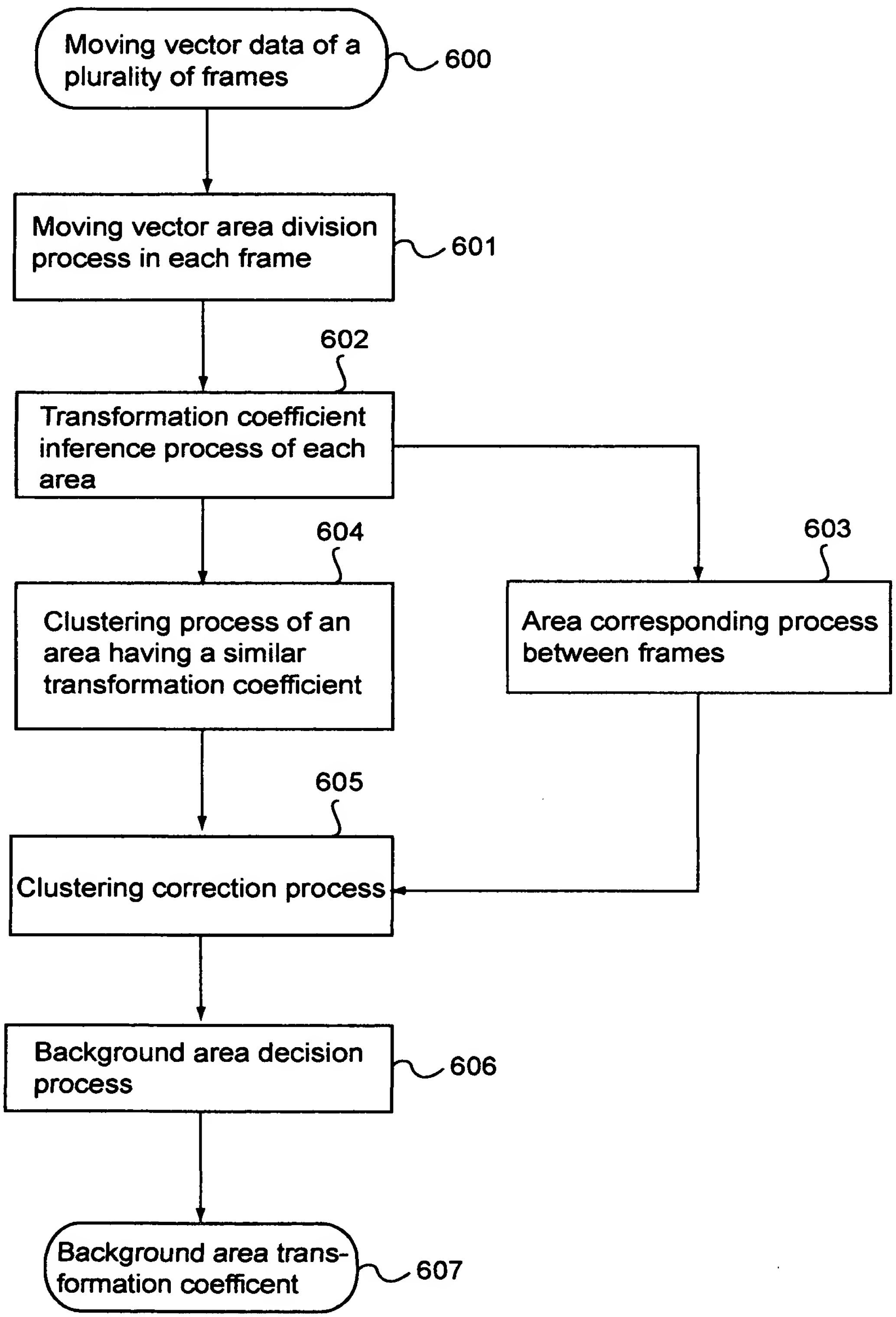

Small dispersal of DC component
→ Low reliable area
Low reliable moving vector
→ Removal
Macro-block

[FIG. 5]

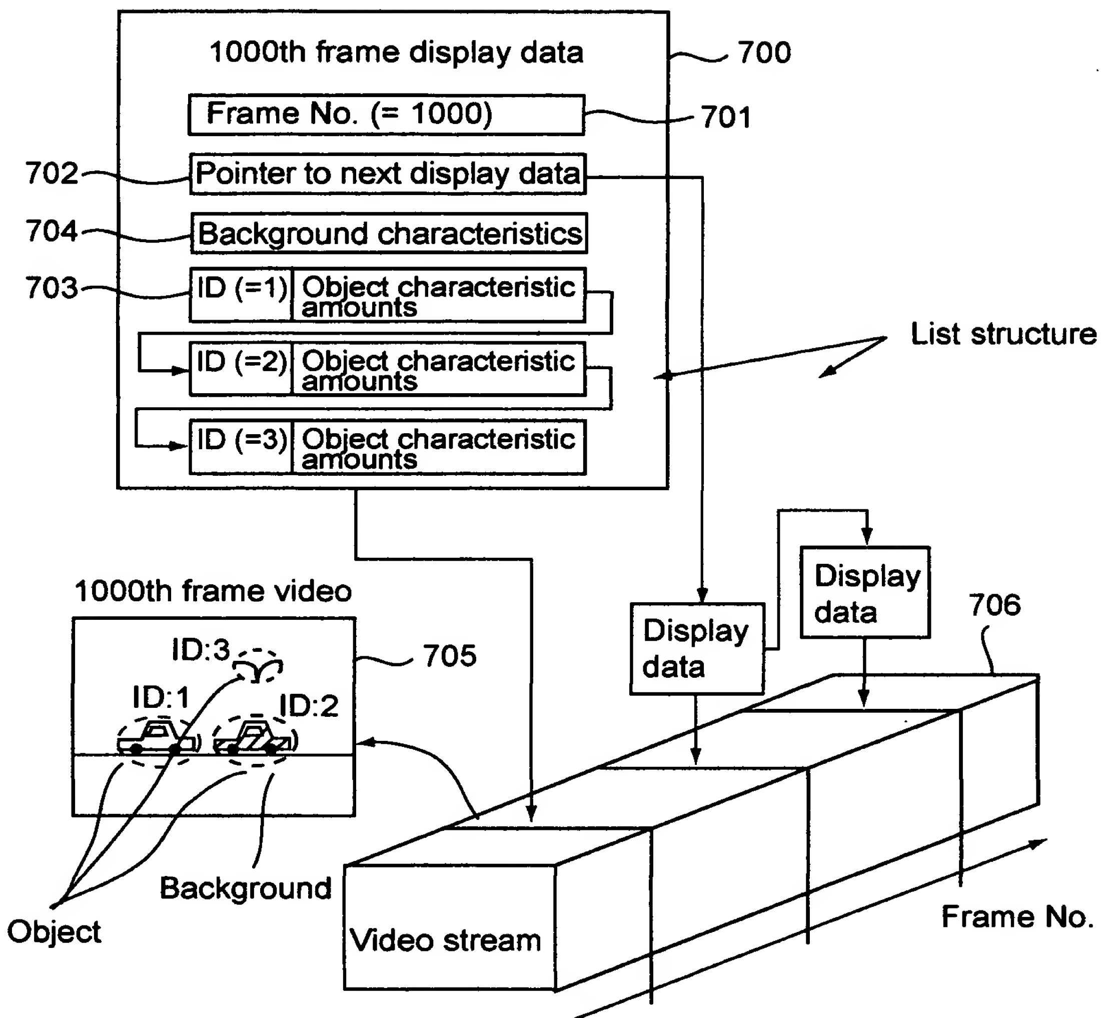


Moving vector data
500
Moving vector area division process
501
Transformation coefficient inference process of each area
502
Clustering process of an area having a similar transformation coefficient
503
Background area decision process
504
Background area trans-formation coefficient
505

[FIG. 6]


Moving vector data of a plurality of frames
600
Moving vector area division process in each frame
601
602
Transformation coefficient inference process of each area
604
603
Clustering process of an area having a similar transformation coefficient
Area corresponding process between frames
605
Clustering correction process
Background area decision process
606
Background area trans-formation coefficent
607

[FIG. 7]



(a)

703

Object characteristic amounts(example)
- Position
- Outline (ellipse, approximate to rectangle)
- Affine tansformation coefficent
- Mean and direction of moving vector
- Color histogram

others

(b)

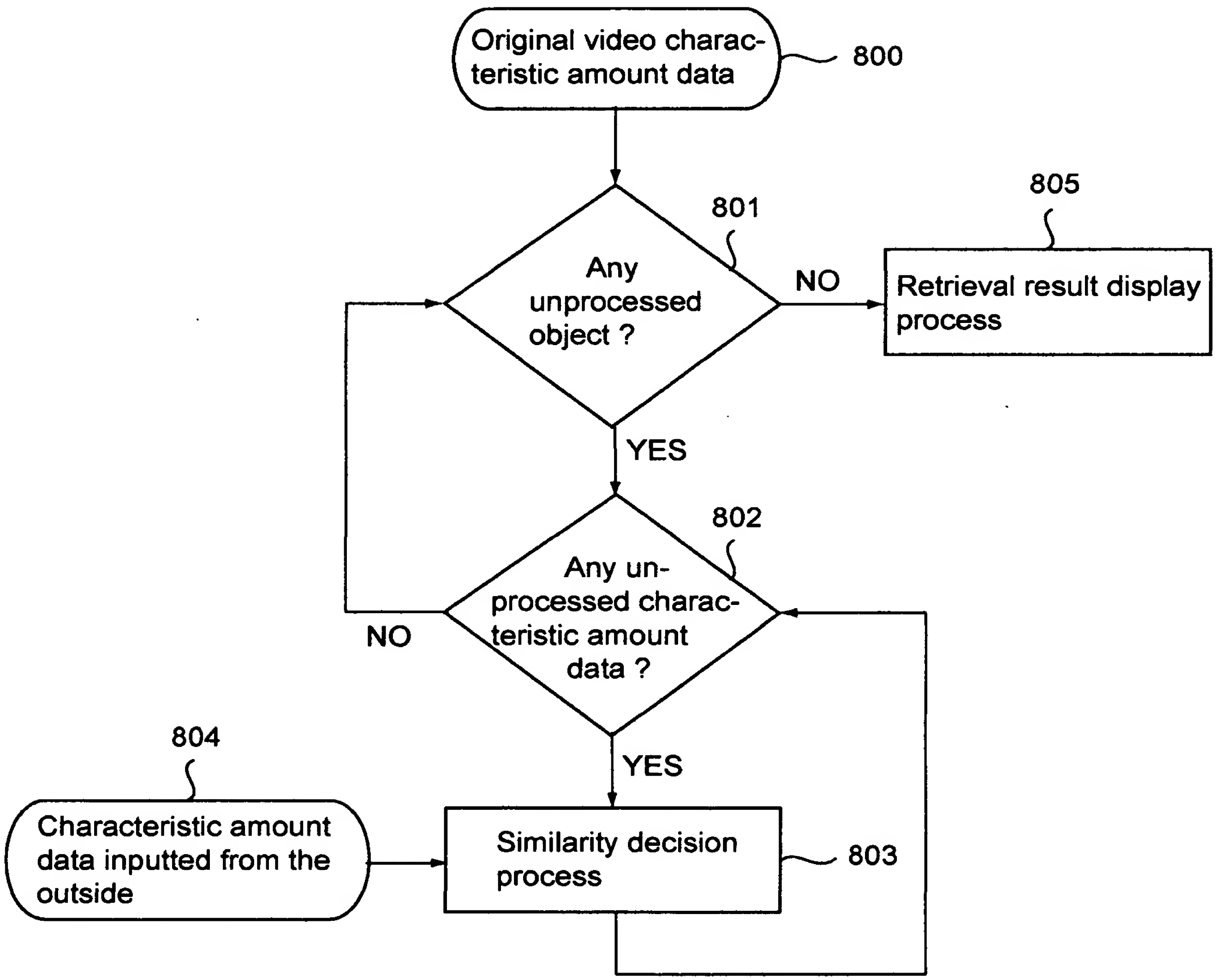
704

Background characteristic amounts (example)
- Affine transformation coeffient
- Mean and direction of moving vector
- Camera work kind
- Color histogram

others

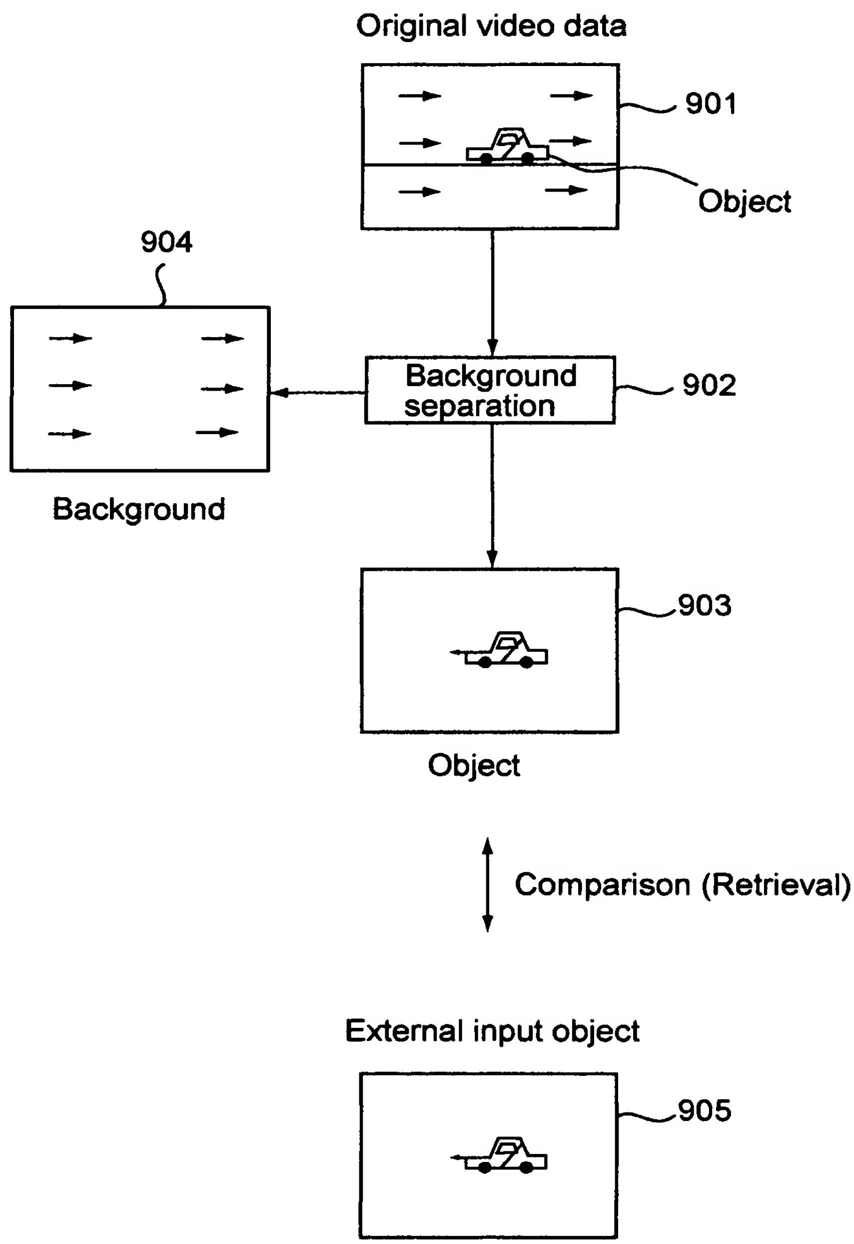
(c)

[FIG. 8]

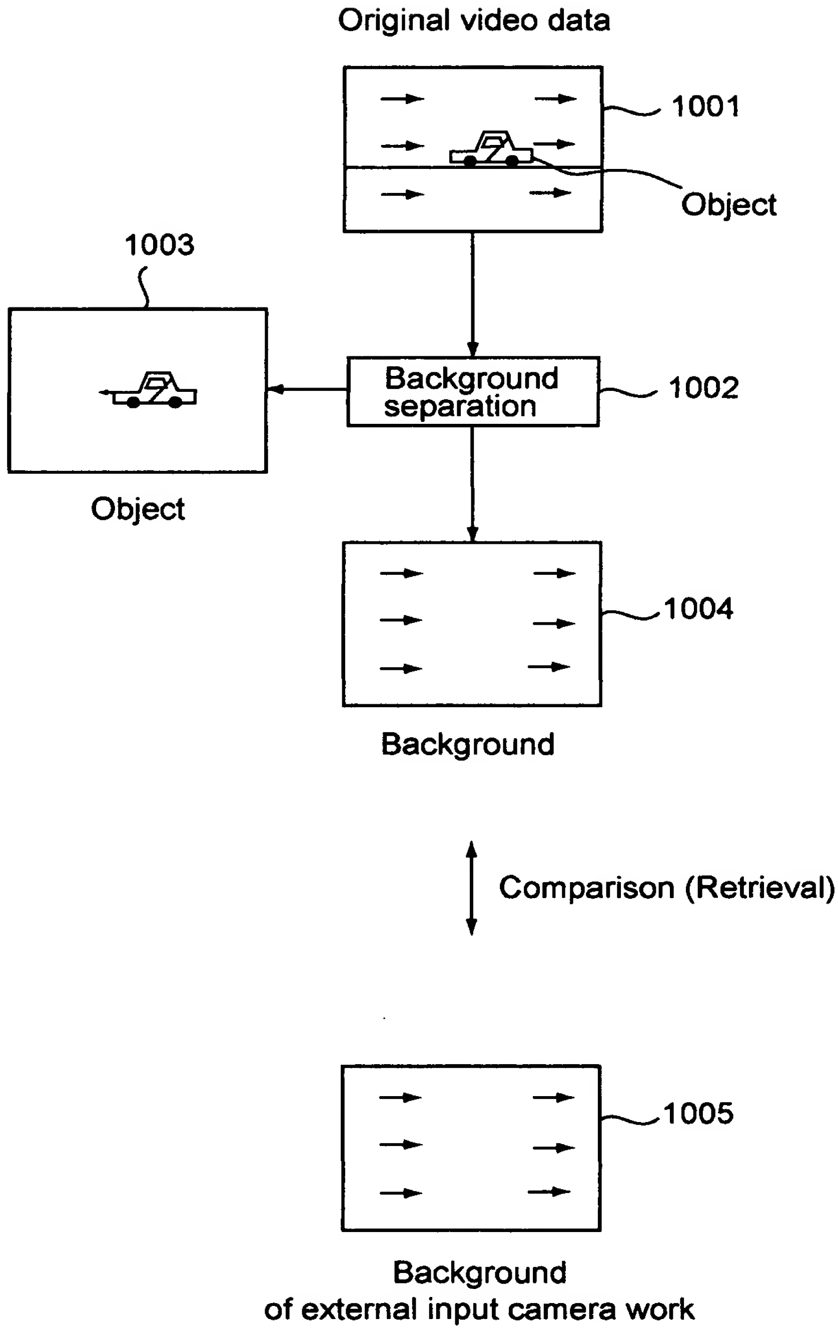


Original video characteristic amount data
800
801
805
Any unprocessed object ?
NO
Retrieval result display process
YES
802
Any unprocessed characteristic amount data ?
NO
YES
804
Characteristic amount data inputted from the outside
Similarity decision process
803

[FIG. 9]

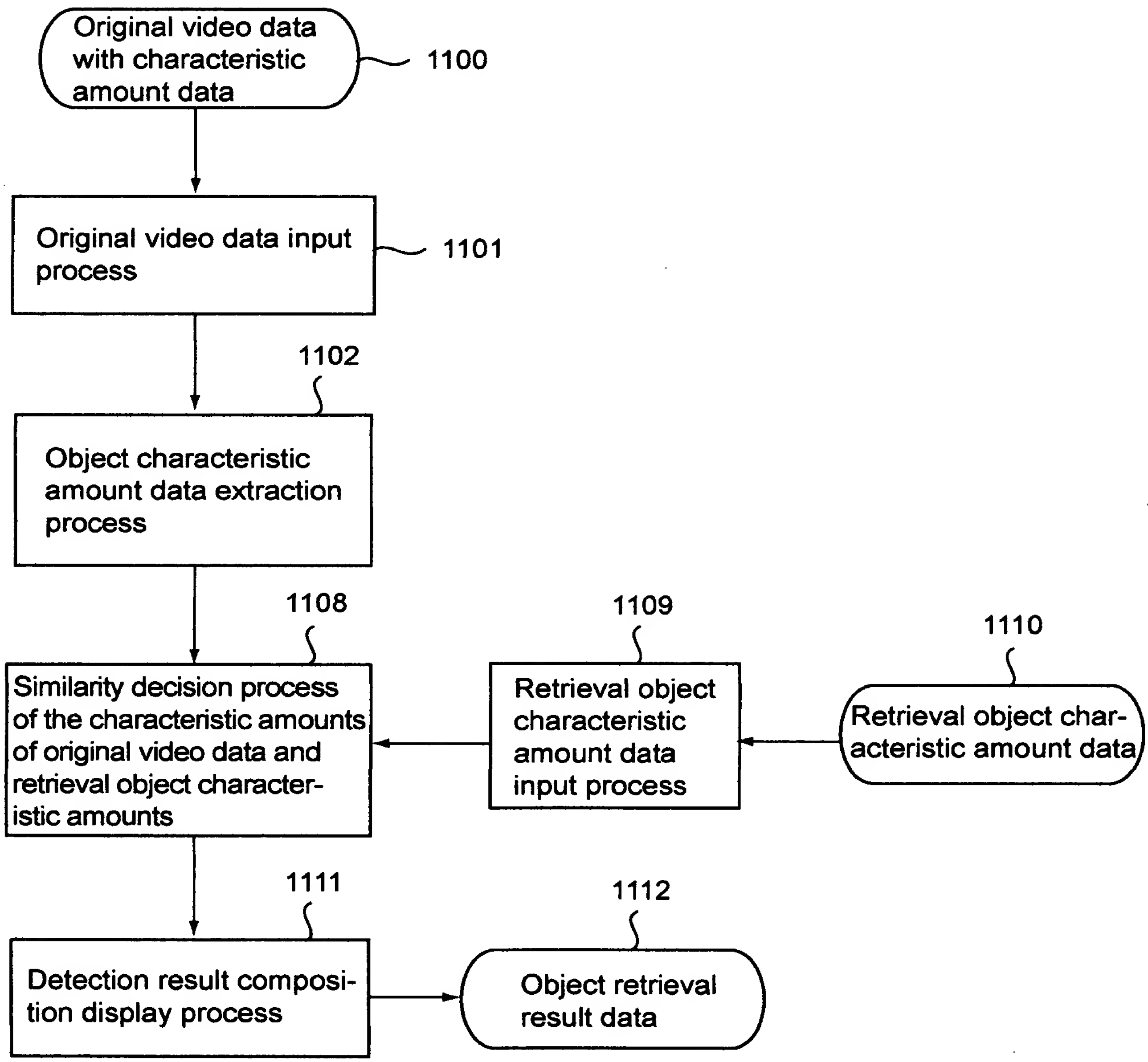

Original video data
901
Object
904
Background separation
902
Background
903
Object
Comparison (Retrieval)
External input object
905

[FIG. 10]



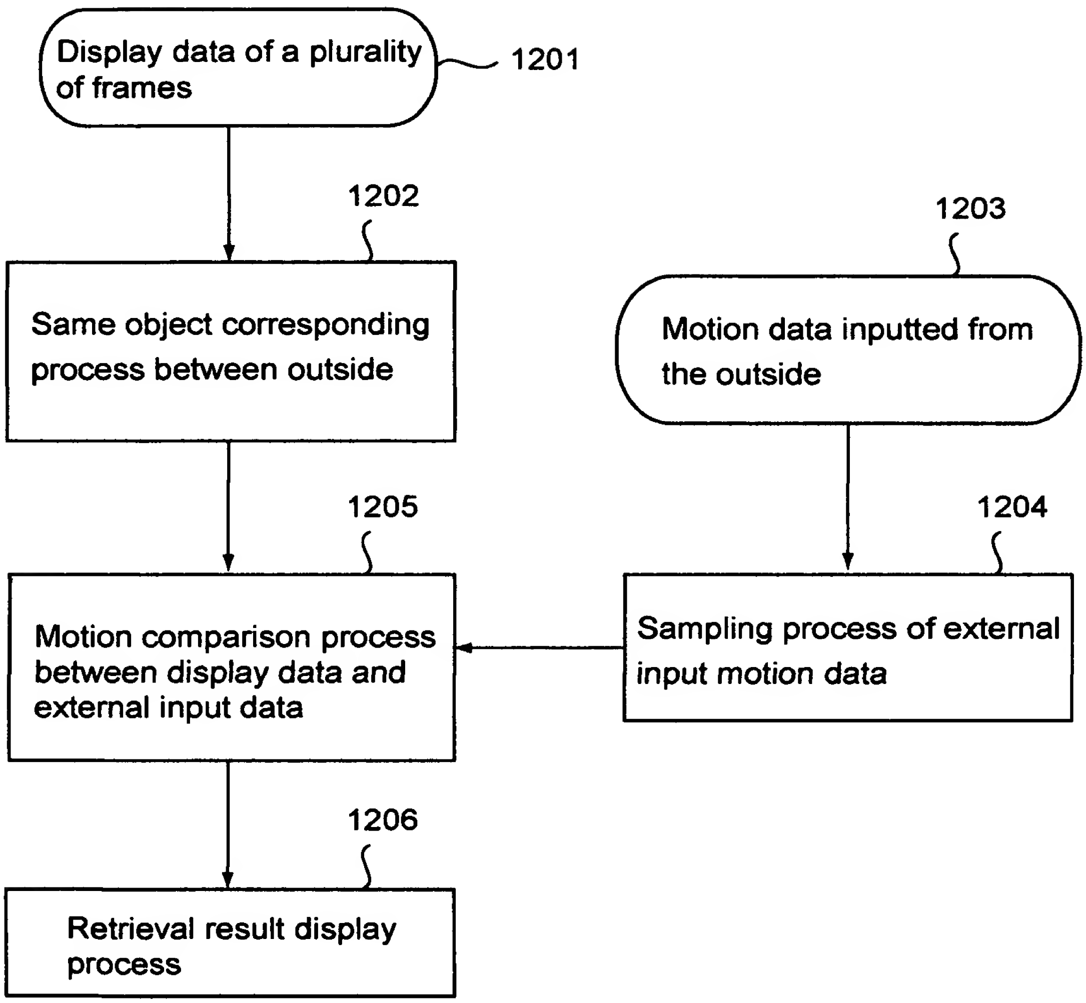
Original video data
1001
Object
1003
Background separation
1002
Object
1004
Background
Comparison (Retrieval)
1005
Background
of external input camera work

[FIG. 11]

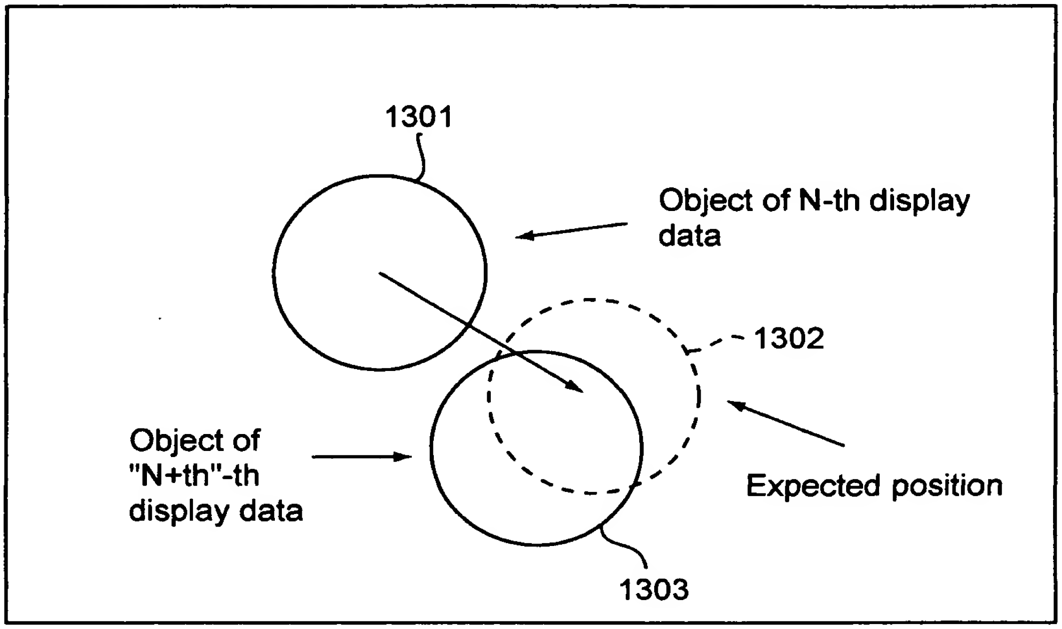


Original video data with characteristic amount data
1100
Original video data input process
1101
1102
Object characteristic amount data extraction process
1108
1109
1110
Similarity decision process of the characteristic amounts of original video data and retrieval object characteristic amounts
Retrieval object characteristic amount data input process
Retrieval object characteristic amount data
1111
1112
Detection result composition display process
Object retrieval result data

[FIG. 12]

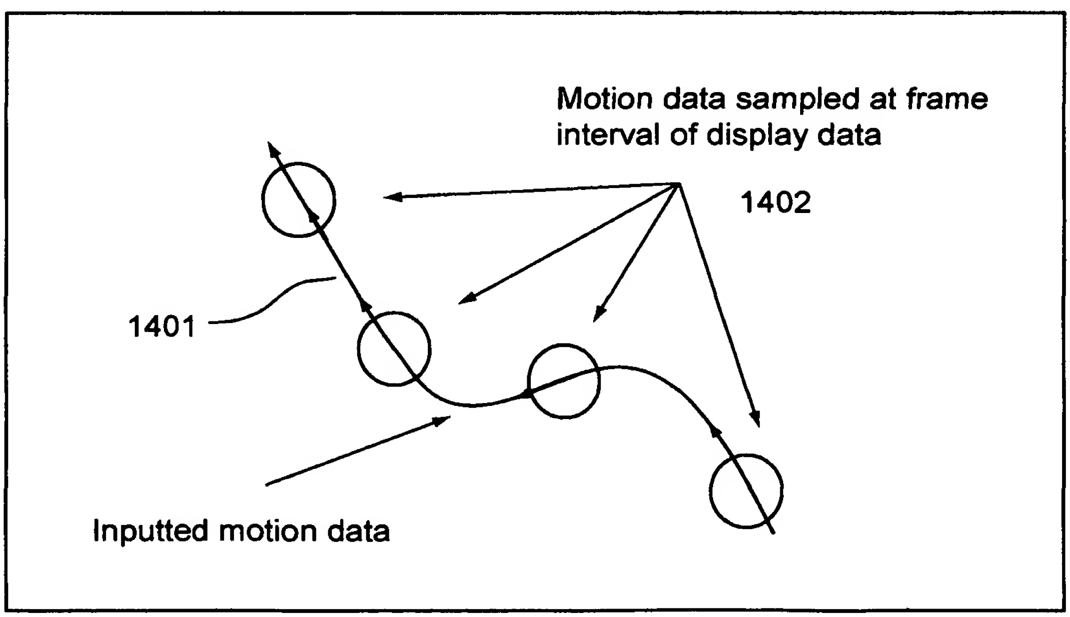

Display data of a plurality of frames
1201
1202
Same object corresponding process between outside
1203
Motion data inputted from the outside
1205
1204
Motion comparison process between display data and external input data
Sampling process of external input motion data
1206
Retrieval result display process

[FIG. 13]



1301
Object of N-th display data
1302
Object of "N+th"-th display data
Expected position
1303

[FIG. 14]



Motion data sampled at frame interval of display data
1402
1401
Inputted motion data